Exhaustive percolation on random networks

Björn Samuelsson* and Joshua E. S. Socolar†

Physics Department and Center for Nonlinear and Complex Systems, Duke University, Durham, NC, 27514

(Dated: February 3, 2008)

We consider propagation models that describe the spreading of an attribute, called "damage", through the nodes of a random network. In some systems, the average fraction of nodes that remain undamaged vanishes in the large system limit, a phenomenon we refer to as *exhaustive percolation.* We derive scaling law exponents and exact results for the distribution of the number of undamaged nodes, valid for a broad class of random networks at the exhaustive percolation transition and in the exhaustive percolation regime. This class includes processes that determine the set of frozen nodes in random Boolean networks with indegree distributions that decay sufficiently rapidly with the number of inputs. Connections between our calculational methods and previous studies of percolation beginning from a single initial node are also pointed out. Central to our approach is the observation that key aspects of damage spreading on a random network are fully characterized by a single function specifying the probability that a given node will be damaged as a function of the fraction of damaged nodes. In addition to our analytical investigations of random networks, we present a numerical example of exhaustive percolation on a directed lattice.

PACS numbers: 89.75.Da, 02.50.Ey, 02.10.Ox, 05.50.+q

I. INTRODUCTION

A. Overview

Propagation models on lattices or more general graphs describe the spreading of some discrete signal through a set of discrete entities. In the most general terms, the signal corresponds to some qualitative change that causes the entity to interact differently with its neighbors. Examples include the spreading of damage in power grids [1, 2], the spreading of disease through a population [3, 4, 5], the spreading of a computer virus on the Internet [6, 7], or the alteration of gene expression patterns in a cell due to a mutation [8, 9]. In the general case, the individual entities are represented as nodes in a graph where the links indicate paths along which the signal can spread [10, 11, 12, 13, 14, 15, 16]. Because the signal can be thought of as disrupting the static or dynamical state of the original system, we refer to its propagation as spreading *damage*, though in many cases the "damage" may enhance a desired property or simply represent some natural dynamical process. A single instance of a given spreading process initiated from a particular subset of nodes is often called an avalanche.

In analyzing spreading processes, one is often interested in the transition between those that die out quickly and those that spread to a finite fraction of the system in the large-system limit, a transition that may occur as the probability of transmitting damage across links is varied. This percolation transition is relevant for systems in which the fraction of initially damaged nodes tends to zero in the limit of infinite system size. The order parameter for the transition is the average fraction of nodes damaged in a single avalanche, which remains zero for small transmission probabilities and continuously increases when the probability rises above a threshold value. We will refer to this as the *sparse percolation* (SP) transition. The SP transition occurs for spreading processes in which the probability that a node becomes damaged is zero unless at least one of its neighbors is damaged. (If this probability were nonzero, a nonzero fraction of the nodes would always get damaged.)

For a certain class of propagation models, there is another transition of interest. When the fraction of initially affected nodes remains fixed as the system size is increased, the fraction of nodes that remain *undamaged* can undergo a transition from finite values to zero at transmission probabilities above some threshold. We refer to this as the *exhaustive percolation* (EP) transition. The EP transition occurs only for propagation models in which the probability of a node remaining undamaged is zero when all of its neighbors are damaged (all of its inputs in the case of a directed graph). We assume also that there is a nonzero probability for a node to remain undamaged if it has at least one undamaged input. There is then one more condition for the EP transition: the density of directed loops of any specified size must vanish in the large system limit. For any loop there is a finite probability that no member of the loop will be damaged, since no member of the loop can have all of its inputs damaged until one of the members becomes damaged through a probabilistic event. Thus EP is *not* observable on spatial lattices of the type generally encountered in statistical mechanics. EP is observable, however, on directed lattices and on graphs in which the nodes serving as inputs to a given node are selected at random.

In this paper we derive the probability distribution for the number of undamaged nodes at the EP transition on random graphs for a general class of propagation mod-

---

*bjorn.samuelsson@duke.edu
†socolar@phy.duke.edu

els exhibiting what we call *unordered binary avalanches* (UBA). This is analogous to finding the distribution of avalanche sizes at the usual percolation transition, but here we are asking for the distribution of the number of nodes *not* participating in the avalanche.

As an application of our EP results, we consider the problem of identifying unfrozen nodes in a random Boolean network (RBN). In a RBN, each node has a binary state that is updated according to a rule that takes the values of some other nodes as inputs. The dynamics of RBNs has been investigated extensively; see, e.g., [17, 18, 19, 20, 21, 22, 23, 24, 25]. A RBN can have several dynamical attractors, but some nodes might have the same value at all times on all attractors. Such nodes are called *stable* and the set of stable nodes is important for the dynamics in RBNs [26, 27, 28].

Almost all stable nodes in a broad class of RBNs can be identified through a dynamic process that was introduced by Flyvbjerg [26] and formalized to facilitate numeric simulations by Bilke and Sjunnesson [27]. We call the stable nodes that can be identified by this dynamic process *frozen* (and nodes that are not frozen are called *unfrozen*). Provided that the Boolean rule distribution is symmetric with respect to inversion of any subset of inputs, the set of frozen nodes can be identified through an UBA in which frozen inputs cause new nodes to become frozen (damaged). Most rule distributions that have been examined in the literature exhibit this symmetry. The requirement is satisfied, for example, for any model that assigns given probability p for obtaining a 1 in each entry of the truth table for each node.

This paper is organized as follows. We first develop the notation and basic definitions required for discussing UBAs in general. In Section I B, we give an introduction to the UBA formalism from the perspective of percolation processes. A more formal description is given in Section II, followed by a numerical illustration of the basic concepts. In Section III, we present analytic derivations for UBA in random networks with emphasis on EP and the EP transition. We also present explicit results for the special case of Erdős–Rényi networks with a natural choice for the avalanche rules.

In Section IV we show how to apply the UBA formalism to obtain the statistics of frozen nodes in two-input RBNs. In the present context, this serves as an illustration of the general theory, but this particular example was also the primary motivation for studying EP. The results on RBNs are consistent with those found by Kaufman, Mihaljev, and Drossel. [29]. The main advantage of using the EP formalism for this problem is that it makes clear how the calculation can be extended to networks with more than two inputs per node, including networks with an in-degree distribution that (with a low probability) allows arbitrarily large in-degrees.

B. Basic definitions

An unordered binary avalanche (UBA) is defined as a spreading process with the following properties:

Binary states: the state of each node can be characterized as a binary variable s, with $s = 0$ meaning *undamaged* and $s = 1$ meaning *damaged*;

Boolean rules: the state of each node is determined by a Boolean function of the states of its input nodes;

Order independence: the probability of having a given set of nodes damaged at the end of the process does not depend upon the order in which nodes are chosen for updating.

Order independence refers to the dynamics of the spreading process or a simulation of it. In such a simulation, one typically chooses a site and updates it according to a rule depending on the states of sites that provide inputs to it, repeating the process until a test of every site yields no change in the state of the system. We are interested in cases where the order in which sites are chosen for possible updating has no bearing on the final state of the system.

UBA is a natural extension of site or bond percolation. To determine the avalanche size distribution in site percolation, for example, one identifies an initial subset of damaged sites and then tests neighbors of damaged sites to see whether the damage spreads to them. After a given site is tested for the first time, its value is permanently fixed. The process is iterated until no new damaged sites are generated. See, e.g., Ref. [30]. This method of investigating site percolation is equivalent to assigning all sites a value, then beginning with a damaged site and determining all of the damaged sites in a connected cluster. Site percolation where each site has the probability p to be occupied can be recast as a UBA system as follows. Let each site be associated with a rule that is an OR-rule of all of its neighbors with probability p and is a constant 0 with probability $1 - p$. Then the above described site percolation is achieved by first selecting the rules and clamping the value of a given site to 1, and then repeatedly updating the system according to the Boolean rules. In this situation, the 1s in the final state mark a site percolation cluster. A more practical way of simulating the same UBA is to determine probabilistically the Boolean rule at each site only when that site is first encountered in the percolation process and to update only those nodes where the rules have been determined.

To ensure order independence in UBA, it is sufficient to require that each Boolean function is non-decreasing, meaning that if one of the inputs to the rule changes from 0 to 1, the output is not allowed to change from 1 to 0. For non-decreasing Boolean functions, if a specific node is eventually going to be assigned the value 1 during an avalanche, updating other nodes to 1 first cannot change the outcome.

We are particularly interested in UBAs that are initiated by damage at a set of nodes comprising a nonzero fraction of the total number of nodes. Such a process would be relevant, for example, if the probability that any given node is damaged at the start is independent of the system size.

To clarify both the distinction between EP (exhaustive percolation) and SP (sparse percolation) and the similarities between them, we describe a particular case of a propagation model that exhibits both transitions. Consider a graph with a total of N nodes, some of which have three input links each while the others have no input links at all. The graph is random in that the node supplying the input value on any given link is selected at random, but stays fixed throughout the avalanche. Let ν_0 be the fraction of nodes with no inputs. Define a spreading process as follows: The initial condition is that all nodes with no inputs are considered damaged. Each other node is now selected in turn to see whether the damage spreads to it. If a node has one damaged input, the probability that it will be damaged is p_1; if it has two damaged inputs, the probability of damage is p_2 (with $p_2 \geq p_1$); and nodes with three damaged inputs are guaranteed to become damaged ($p_3 = 1$). These probabilities are realized, for example, by the following Boolean rule distribution: a 3-input OR-rule with probability p_1; a 3-input majority rule with probability $p_2 - p_1$; and a 3-input AND-rule with probability $1 - p_2$.

As N goes to infinity, the number of initially damaged nodes can be a nonzero number that grows slower than N, meaning that ν_0 goes to zero as N goes to infinity. In this limit, the SP transition occurs at $p_1 = 1/3$ and the spreading from each initially damaged node is described by a *Galton–Watson process*. In a Galton–Watson process, a tree is created by adding branches to existing nodes, with the number of branches emerging from each node drawn from a fixed probability distribution. Such branching processes have been investigated extensively. (See, e.g., Ref. [31].) In particular, the correspondence to Galton–Watson processes means that for critical SP, the probability of finding n damaged nodes scales like $n^{-3/2}$ for $1 \ll n \ll N$ [9, 32].

For any nonzero value of ν_0, the EP transition occurs for p_2 satisfying $(1 - p_2)(1 - \nu_0) = 1/3$ (assuming that this value of p_2 is greater than p_1.) The analysis described in Section III provides a method of calculating the probability $P(u)$ of having u *undamaged* nodes in this case. The result in the large N limit is $P(u) \sim P(0)u^{-1/2}$ for large u. A difference between EP and SP is that both $P(0)$ and the cutoff on the $u^{-1/2}$ distribution scale with N for EP, while for SP only the cutoff scales with N.

II. INTRODUCTION TO EXHAUSTIVE PERCOLATION

A. Formal description of UBA

We now describe a formalism and establish some notation that is suitable for a detailed treatment of UBA. Let N denote the number of nodes in a network with a specified set of links and let the nodes be indexed by $j = 1, \ldots, N$. The network state is described by the vector $\mathbf{s} = \{s_1, \ldots, s_N\}$. Let K_j denote the number of inputs to node j, and let $\mathbf{k}_j$ denote the vector of K_j inputs to node j. Furthermore, let R denote a Boolean function and let $\Pi_j(R)$ denote the initial probability that node j has the rule R. [It is required that R has precisely K_j inputs for $\Pi_j(R)$ to be nonzero.]

To efficiently simulate UBA, we keep track of the information that is known about each node at each step in the process. In particular, it is important to keep track of whether or not the change from 0 to 1 of a given input has already been accounted for in determining the output. The simplest way to do this is to introduce an extra state 0* that labels a site whose rule R implies an output value of 1 but for which the update to 1 has not yet been implemented. When a node changes its state from 0 to 0*, it is a silent change in the sense that the Boolean rules at the other nodes treat an input 0* exactly the same as 0. To retrieve the final state of the network, all occurrences of 0* must be updated to 1. When a single update to 1 is made, the information that the given node has value 1 is passed along to all nodes with inputs from it. The values of these nodes may then change from 0 to 0*. The conditional probability that the value of node i is updated from 0 to 0* when j changes value from 0* to 1, is given by

$$U_i(\mathbf{s}, j) \equiv \frac{P_1(\mathbf{k}_i') - P_1(\mathbf{k}_i)}{1 - P_1(\mathbf{k}_i)}, \qquad (1)$$

where $\mathbf{k}_i'$ is the value of $\mathbf{k}_i$ after s_j has been updated and $P_1(\mathbf{k}_i)$ is the probability that $R_i(\mathbf{k}_i) = 1$:

$$P_1(\mathbf{k}_i) \equiv \sum_R R(\mathbf{k}_i)\Pi_i(R). \qquad (2)$$

The numerator in Eq. (1) is the probability that R_i produces a 1 after the update of node j minus the probability that R_i produced a 1 before the update. The denominator is the probability that node i had the value 0 before the update.

Let $\Pi_i(1)$ denote the probability that the rule at node i has output 1 regardless of its input values. If some particular nodes are selected for initiation of the UBA, $\Pi_i(1)$ is set to one for these nodes [which means $\Pi_i(R) = 0$ for all other rules].

We are now ready to present a formal algorithm for determining the final state of an instance of UBA on a finite network. We carry out the following procedure (where := denotes the assignment operator):

1. $s_j := 0$ for all j;

2. $s_j := 0^*$ with probability $\Pi_j(1)$ for each j;

3. Some j with $s_j = 0^*$ is selected;

4. $s_i := 0^*$ with probability $U_i(\mathbf{s}, j)$ for each i with $s_i = 0$;

5. $s_j := 1$;

6. Steps 3–5 are iterated as long as there exists a node in state 0^*.

UBA can also be considered on infinite networks, but that requires a more technical description of the process. First, the choices of j in step 3 for both descriptions must be such that any given j that satisfies the conditions in step 3 will be selected in a finite number of iterations. Second, the ensemble of final states needs to be defined in terms of a suitable limit process because the stopping criterion in step 6 can not be applied to an infinite system.

Note that the dynamics is only dependent on the probability functions $\{P_1(\mathbf{k}_i)\}$. That is, the precise rule distributions affect the avalanche results only through their contributions to P_1. Because the Boolean rules are non-decreasing functions, $P_1(\mathbf{k}_i)$ is also a non-decreasing function. In fact, every non-decreasing function, $f(\mathbf{k}_i)$, with values in the interval $[0, 1]$ can be realized by $P_1(\mathbf{k}_i)$ for a suitable Boolean rule distribution. One such rule distribution can be constructed as follow: for each i and each $\mathbf{k}_i$, select a random number y from a uniform distribution on the unit interval and set $R_i(\mathbf{k}_i) = 1$ if and only if $y < f(\mathbf{k}_i)$.

B. An example of EP on a lattice

To illustrate the concepts of UBA and EP, consider a directed network on a two-dimensional square lattice with periodic boundary conditions. Each node in the lattice has integral coordinates (i, j) where $i + j$ is odd and the node at (i, j) receives inputs from the two nodes at $(i-1, j\pm1)$. The rule for propagation of damage to a node is either OR or AND, with probabilities $\Pi_{(i,j)}(\text{OR}) = r$ and $\Pi_{(i,j)}(\text{AND}) = 1 - r$, respectively.

Figure 1 displays an avalanche that is initiated by letting each node be initially damaged with probability $\rho = 1/8$. A node assigned OR becomes damaged if either of its neighbors one layer above is damaged; a node assigned AND becomes damaged if and only if both neighbors above it are damaged. This means that

$$P_1(\mathbf{k}_{(i,j)}) = \begin{cases} 0 & \text{if } \mathbf{k}_{(i,j)} = (0,0) \\ r & \text{if } \mathbf{k}_{(i,j)} \in \{(0,1), (1,0)\} \\ 1 & \text{if } \mathbf{k}_{(i,j)} = (1,1) . \end{cases} \quad (3)$$

Note that clusters of damaged nodes formed in an avalanche initiated by a single damaged node cannot contain any holes, as the uppermost undamaged node in the
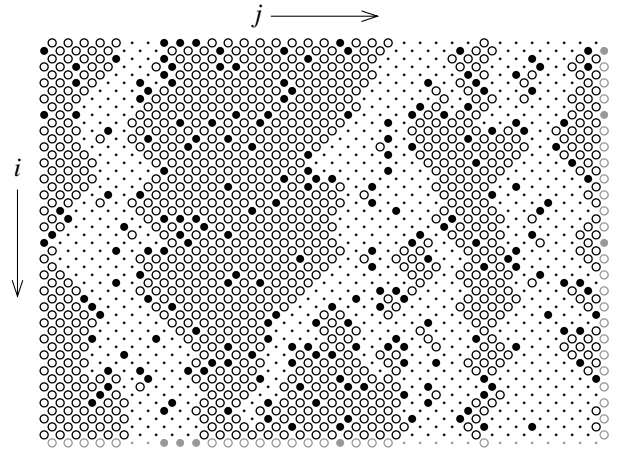

FIG. 1: An example of UBA on a lattice, displaying undamaged nodes (dots), initially damaged nodes (filled circles), and nodes damaged during the avalanche (empty circles). Each node has either an OR-rule or an AND-rule with inputs from its neighbors in the row immediately above the node. The probability for a node to be initially damaged is $\rho = 1/8$ and the probability for obtaining an OR-rule is $r = 0.3$. Periodic boundary conditions are used and the first row and column are repeated in gray after the last row and column to illustrate the periodic boundary conditions.

hole would have to have two damaged inputs and hence would become damaged when updated.

For localized initial damage, the SP threshold is found at $r = 1/2$. Above this value of r, domains of damage tend to widen as the avalanche proceeds. Since the growing cluster has no holes, this is simultaneously an EP transition. The EP transition can be found for smaller values of r in lattices where each node is initially damaged with a given nonzero probability ρ. [For every initially damaged node, $\Pi_{(i,j)}(1)$ is set to 1, meaning that $P_1(\mathbf{k}_{(i,j)}) = 1$ for every value of $\mathbf{k}_{(i,j)}$.]

Figure 2 shows the average number of unaffected nodes as a function of r for $\rho = 1/8$ on lattices with periodic boundary conditions. The numerics displayed in Fig. 2 clearly suggest that there is a second-order EP phase transition. Furthermore, these numerical results suggest that the avalanche in Fig. 1 is within the parameter regime for EP and that EP does not occur in this case due to finite size effects.

For the case $r = 0$, it is possible to map the EP transition onto ordinary, directed, site percolation on the same lattice. When all nodes in the lattice have AND-rules, the following algorithm may be used to determine whether a given node will be damaged: select a node; put a mark on the selected node unless it is initially damaged; and recursively mark each initially undamaged node that has an output to a marked node. The selected node will get damaged if and only if this recursion ends in a finite number of steps. The algorithm describes ordinary
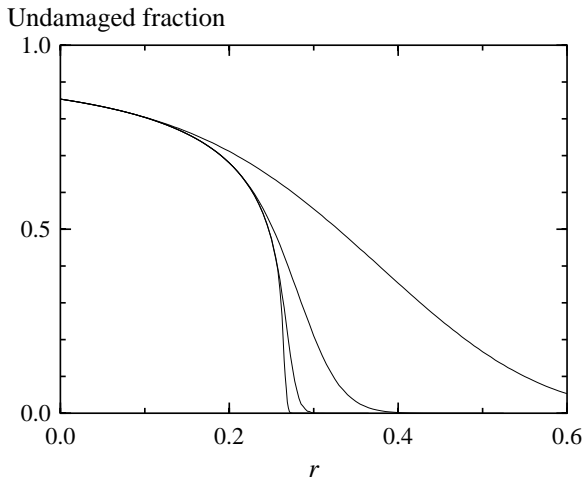
Undamaged fraction

FIG. 2: The average fraction of undamaged nodes for UBA on a lattice of the type shown in Fig. 1, as a function of the selection probability p for OR-rules and the probability $\rho = 1/8$ for initial damage. The lattice has periodic boundary conditions and covers a square that has a side of 10, 10^2, 10^3, and 10^4 lattice points, respectively, with steeper curves for larger systems. The statistical uncertainty in the estimated mean is less than the line width.

directed site percolation where the initially undamaged (damaged) nodes are considered active (inactive) sites and the process propagates in the opposite direction relative to the UBA. We therefore expect the EP transition to occur for a value of ρ equal to $1 - p_c$, where $p_c = 0.70549$ is the threshold for directed site percolation [33] and we have confirmed this with numerical tests. Further study of EP on the lattice is beyond the scope of this paper.

C. Suppression of EP by resistant motifs

In the lattice example above, the fact that the network had no feedback loops smaller than the lattice size was important. In general, EP is suppressed by the presence of short feedback loops. As already noted, for EP to occur, it is required that the output of each rule in the rule distribution is 1 if all of its inputs have the value 1. Otherwise, there would be a finite fraction of nodes that keep the value 0 regardless of the influence from the rest of the network. Generalization of this reasoning allows us to rule out EP in other situations, indicating that EP is most likely to occur in directed or highly disordered networks. To pursue this idea, we introduce the notion of *resistant motifs*.

A *motif* is a small network with a particular arrangement of internal links. A given motif may occur many times in a network with different rules assigned to its nodes and with different configurations of external inputs. A motif is *resistant* with respect to a given ensemble of rule assignments if the probability of damage

entering the motif when all external inputs are damaged is strictly less than unity. For the rule distributions that we consider for the EP transition in random networks, each node has a nonzero probability of being assigned a rule that sets its output to 0 if at least one of its inputs is 0. Thus when all of the nodes in a feedback loop of any length have the value 0, there is a nonzero probability that they will all remain 0 even if all external inputs to the loop are set to 1. Every feedback loop of a given length is therefore a resistant motif.

If the number of occurrences of a resistant motif grows linearly with the network size, there will in total be a finite fraction of nodes that remain unaffected with a finite probability. For such networks, EP cannot occur in the limit of large systems. Examples include typically studied regular lattices and small world networks with link directions assigned so that short feedback loops are prevalent.

The problem resistant motifs can be avoided in random networks having a mean indegree $\langle K \rangle$ that is well-defined and independent of N, in which case the number of feedback loops of a given length approaches a constant. Though the total number of resistant motifs may grow with system size, the larger motifs have a low probability of avoiding damage. For large N, the out-degree distribution is a Poisson distribution with a mean value of $\langle K \rangle$. The outputs emerging from a given node form a tree with approximately $\langle K \rangle^m$ nodes at the mth level. Thus, the probability for a given node to be part of a cycle of m nodes is approximately $\langle K \rangle^m / N$, which means that the typical number of feedback loops of length m is approximately $\langle K \rangle^m / m$. On the other hand, the loop may contain either initially damaged nodes or some rules that allow damage to enter from external inputs. The probability that this will *not* occur decays exponentially with m. If the decay is faster than $\langle K \rangle^{-m}$, the density of nodes in undamaged resistant motifs will approach zero.

In summary, EP (for the considered type of rule distributions) is excluded on lattices with a high density of feedback loops. For random networks, however, the fraction of nodes in undamaged resistant motifs can go to zero in the large N limit. This property allows EP to occur on random networks as demonstrated in the following section.

III. EP ON RANDOM NETWORKS

A. Criteria for EP

Consider a network such that the inputs to each node are chosen randomly and uniformly from all nodes in the network and the probability functions $\{P_1(\mathbf{k}_i)\}$ are determined from a given distribution of Boolean rules. For such networks, UBA can be handled analytically.

Define $g(x)$ as the probability for a rule in the random network to output 1 if each input has the value 1 with probability x. The function g reflects the probabil-

ity for propagation of damage to a single node, for the considered instance of UBA. We refer to g as the *damage propagation function*. In random networks, $P_1(\mathbf{k}_i)$ is independent of i and can be replaced by $P_1(\mathbf{k})$. Let K denote the number of components of $\mathbf{k}$, i.e., the number of inputs to the considered node. $g(x)$ can then be expressed as

$$g(x) = \sum_{K=0}^{\infty} P(K) \sum_{\mathbf{k} \in \{0,1\}^K} x^I (1-x)^{K-I} P_1(\mathbf{k}), \quad (4)$$

where I is the number of 1s in $\mathbf{k}$ and $P(K)$ is the probability to draw a rule with K inputs.

Let N denote the total number of nodes, and let n_0, n_{0*}, and n_1 denote the number of nodes with the values 0, 0*, and 1, respectively. With these definitions and the fact that $U_i(\mathbf{s}, j)$ is independent of i for the random network, the role of $\{P_1(\mathbf{k}_i)\}$ is taken over by $g(n_1/N)$ and Eq. (1) is replaced by

$$U(\mathbf{s}, j) = \frac{g(n_1'/N) - g(n_1/N)}{1 - g(n_1/N)}, \quad (5)$$

where $n_1' = n_1 + 1$. This means that the size of the network, the number of initially damaged nodes, and the damage propagation function g taken together are sufficient to uniquely determine the stochastic spreading process.

After one pass of the update steps 3–5 (from Section II A), the new values n_0' and n_{0*}' of n_0 and n_{0*} are given by

$$n_0' = n_0 - \delta \quad (6)$$

and

$$n_{0*}' = n_{0*} + \delta - 1 \quad (7)$$

where

$$\delta = B_{n_0}[U(\mathbf{s}, j)], \quad (8)$$

with $B_n(a)$ being a stochastic function that returns the number of selected items among n items if the selection probability for each of them is a. The avalanche ends when $n_{0*} = 0$.

The number of damaged nodes, n, in a complete avalanche is the final value of n_1, whereas the number of undamaged nodes, u, is the final value of n_0. An order parameter for the system is $\phi = \lim_{N \to \infty} \langle n/N \rangle$, where the average is taken over the ensemble of networks. The SP transition is found when ϕ changes from zero to a nonzero value, whereas the EP transition is found when ϕ reaches 1.

To understand the typical development of an avalanche, it is convenient to change from the variables n_0, n_{0*}, and n_1, which are constrained to sum to N, to the variables $x_1 \equiv n_1/N$ and

$$c \equiv \frac{n_0}{1 - g(x_1)}. \quad (9)$$

As long as $n_{0*} > 0$, the average value of c after a single update is given by

$$\langle c' \rangle = \frac{\langle n_0' \rangle}{1 - g(x_1')} \quad (10)$$

$$= \frac{n_0 - c[g(x_1') - g(x_1)]}{1 - g(x_1')} \quad (11)$$

$$= c. \quad (12)$$

Hence, as long as $n_{0*} > 0$ for all members of an ensemble of avalanches, $\langle c \rangle$ (the average of c over the ensemble) is conserved as the avalanche proceeds.

From Eqs. (5)–(8) and the definition of c, the variance in c can be calculated. We begin by computing the increment of the variance due to one update step, $\sigma^2(c')$. To leading order as $N \to \infty$, we get

$$\sigma^2(c') = \frac{\sigma^2(\delta_0)}{[1 - g(x_1')]^2} \quad (13)$$

$$= \frac{n_0 U(\mathbf{s}, j)[1 - U(\mathbf{s}, j)]}{[1 - g(x_1')]^2} \quad (14)$$

$$= \frac{c\, U(\mathbf{s}, j)}{1 - g(x_1)} \quad (15)$$

$$= \frac{c}{N[1 - g(x_1)]^2} \left. \frac{dg(x)}{dx} \right|_{x=x_1}. \quad (16)$$

Eq. (16) gives the increment of the variance of c from one update step. To get the total variance of c, we need to sum over all updates from $n_1 = 0$ to the desired value of n_1. Provided that there is an upper bound κ such that $dg(x)/dx < \kappa$ for all x, the total variance of c satisfies

$$\sigma_{\text{tot}}^2(c) < n_1 \frac{c\,\kappa}{N[1 - g(x_1)]^2} < \frac{\kappa N}{1 - g(x_1)} \quad (17)$$

for $x_1 < 1$. (Note that $1/[1 - g(x)]$ is a nondecreasing function because $g(x)$ is nondecreasing.)

The avalanche is initiated with $n_{0*} \equiv n_{0*}^{\text{i}}$, $n_0 = N - n_{0*}^{\text{i}}$, and $n_1 = 0$. The process ends when $n_0 + n_1 = N$ and we seek the distribution of n_0 or n_1 when this happens. According to Eq. (17), the standard deviation of c/N scales like $1/\sqrt{N}$, which implies that both n_0/N and n_{0*}/N have standard deviations that scale like $1/\sqrt{N}$. (x_1 has zero standard deviation because n_1 is incremented by exactly unity on every update step.) Thus in the large system limit, the probability of any member of the the ensemble of avalanches stopping is negligibly small as long as n_{0*}/N is finite, and we may treat c as exactly conserved as long as this condition holds.

Using the initial values $x_1 = 0$ and $n_0 = N - n_{0*}^{\text{i}}$, which determine c, Eq. (9) can be rearranged to give

$$n_0 = [1 - g(x_1)] \frac{N - n_{0*}^{\text{i}}}{1 - g(0)}. \quad (18)$$

Noting that $n_0/N = 1 - x_1 - n_{0*}/N$, we see that in the large N limit, the process continues as long as the strict

inequality

$$1 - x_1 > [1 - g(x_1)] \frac{1 - \lim\limits_{N \to \infty} n_{0*}^{\mathrm{i}}/N}{1 - g(0)} \qquad (19)$$

holds, since the inequality implies that n_{0*}/N remains finite. Moreover, in the large N limit it is impossible to reach values of x_1 for which the inequality has the opposite sign, because the process stops when n_{0*} reaches zero.

Note that because of the zero probability of a node remaining undamaged when all of its neighbors are damaged, we have $g(1) = 1$, which in turn implies that Eq. (19) becomes an equality at $x_1 = 1$. If Eq. (19) is satisfied for all $x_1 < 1$, the process will be exhaustive in the sense that it will not end with a finite value of n_0/N. If, on the other hand, the inequality changes sign for x_1 above some threshold value, then the process will terminate when the threshold is reached. If the left hand side of Eq. (19) forms a tangent line to the right hand side of the expression at some value of x_1, the process will exhibit critical scaling laws. The critical case for EP occurs when the when the tangency occurs at $x_1 = 1$. Examples of these behaviors are presented below and in Section IV.

As an aside, we note that the SP transition is an instance of criticality at $x_1 = 0$. For the above mentioned criterion of criticality to hold at $x_1 = 0$, the right hand side of Eq. (19) must have the value 1 and the slope -1 at $x_1 = 0$. Thus, the system is critical with respect to SP if $\lim_{N \to \infty} n_{0*}^{\mathrm{i}}/N = 0$ and

$$\left. \frac{dg(x)}{dx} \right|_{x=0} = 1 - g(0). \qquad (20)$$

Eqs. (4) and (20) immediately give a criterion for critical percolation on graphs in which every possible directed link (including self-inputs) exists with an independent, fixed probability, assuming the conventional choice in which damage spreads to a given node with probability p from each of its damaged neighbors. In this case we have

$$g(x) = \sum_{K=0}^{\infty} P(K) \left[1 - (1 - px)^K \right], \qquad (21)$$

which yields

$$\left. \frac{dg(x)}{dx} \right|_{x=0} = p \sum_{K=0}^{\infty} P(K) K \qquad (22)$$

$$= p \langle K \rangle. \qquad (23)$$

This result is closely related to the well-known criterion for the presence of a percolating cluster in an Erdős–Rényi graph: percolation occurs when the probability p_{ER} for the presence of a link between two randomly selected nodes exceeds $1/N$, where N is the number of nodes. [34] In the present context, p_{ER} is mapped to $p_{\mathrm{link}}p$, where

p_{link} is the probability that a link exists connecting the two randomly selected nodes and p is the probability that damage spreads across that link. At the same time, we have $\langle K \rangle = p_{\mathrm{link}}N$. (Recall that K is only the indegree of a node, not the total number of links connected to it.) Thus Eq. (23), which implies that the critical value of p is $1/\langle K \rangle$, is consistent with the well-known theory of Erdős–Rényi graphs. [34]

Eq. (23) applies for any distribution of indegrees so long as $\langle K \rangle$ is well-defined and the source of each input is selected at random (so that the outdegrees are Poisson distributed). We note that the latter condition is *not* met by random regular graphs (graphs in which all nodes have the same outdegree) because the probabilities of two nodes getting an output from the same node are correlated.

SP can also be understood by the theory of Galton–Watson processes. If $\lim_{N \to \infty} n_{0*}^{\mathrm{i}}/N = 0$, the update described by Eqs. (5)–(8) is consistent with a Galton–Watson processes that has a Poisson out-degree distribution with a mean value

$$\lambda = \frac{1}{1 - g(0)} \left. \frac{dg(x)}{dx} \right|_{x=0}. \qquad (24)$$

See References [9, 31, 32]. See Appendix C for more details on SP in relation to known results. Cases of tangencies at intermediate values of x_1 are beyond the scope of the present work.

Returning to the question of the EP transition, it is convenient to change variables once again. We define $x_{0,0*} \equiv 1 - x_1$ and $q(x_{0,0*}) \equiv 1 - g(x_1)$. In words, $q(x)$ is the probability that a randomly selected node will output 0 given that each of its inputs has the value 0 with probability x. We refer to q as the *damage control function* as it characterizes the probability that damage will be prevented from spreading to a single node. Equation (19) is then transformed to

$$x_{0,0*} > q(x_{0,0*}) \frac{1 - \lim\limits_{N \to \infty} n_{0*}^{\mathrm{i}}/N}{q(1)}. \qquad (25)$$

Critical EP is found when the left hand side of Eq. (25) forms a tangent line to the right hand side of the expression at $x_{0,0*} = 0$. At criticality, the right hand side of Eq. (25) should have the value 0 and the slope 1. Hence, the conditions $q(0) = 0$ and

$$\left. \frac{dq(x)}{dx} \right|_{x=0} = \frac{q(1)}{1 - n_{0*}^{\mathrm{i}}/N} \qquad (26)$$

are required for an EP transition.

Example: EP on random digraphs

We now consider the special case of graphs in which every possible directed link (including self-inputs) exists with an independent, fixed probability. (We have already

discussed SP on such graphs.) If damage spreads along each directed link with probability p, there is no EP transition because there is a nonzero probability for a node to remain undamaged when all of its inputs are damaged. A minimal change that allows EP on such graphs is to give a special treatment to nodes whose inputs are all damaged, in which case the considered node should always get damaged. For the same reason, all nodes with no inputs must be initially damaged. Other nodes might also be initially damaged, and we let this happen with a given probability ρ for each node with at least one input. For such a network, we can calculate the damage propagation function according to

$$g(x) = \sum_{K=0}^{\infty} P(K) \big[1 - (1-px)^K + (1-p)^K x^K \big] \quad (27)$$

$$= 1 - e^{-\langle K \rangle px} \big(1 - e^{-\langle K \rangle (1-x)} \big). \quad (28)$$

The corresponding damage control function becomes

$$q(x) = e^{-\langle K \rangle p(1-x)} \big(1 - e^{-\langle K \rangle x} \big). \quad (29)$$

A necessary condition for the EP transition is derived from Eq. (26), yielding

$$\langle K \rangle e^{-p \langle K \rangle} = \frac{1}{1-\rho} . \quad (30)$$

For the EP transition to occur, it is also required that

$$f(x) \equiv x - q(x)(1-\rho) \geq 0 \quad (31)$$

for all $x \in [0,1]$ according to Eq. (25). If both Eqs. (30) and (31) are satisfied, the EP transition occurs at the value of p given by Eq. (30):

$$p_c = \frac{\ln \langle K \rangle + \ln(1-\rho)}{\langle K \rangle} . \quad (32)$$

Equation (30) turns out to be a sufficient and necessary condition for the EP transition. Provided that Eq. (30) holds, the first derivative satisfies $f'(0) = 0$. From the observation $f'''(x) < 0$, it is then straightforward to show that $f(x)$ has no local minimum on the interval $(0,1)$. Since $f(0) = 0$ and $f(1) > 0$, Eq. (31) holds for all $x \in [0,1]$.

It is instructive to examine the phase diagram at fixed ρ. A negative value of p_c indicates that the system is always in the EP regime, so for $\langle K \rangle < 1$ the system exhibits EP and it is not possible to observe a transition. For $\langle K \rangle > 1$, an EP transition can be observed at $p = p_c$. A curious feature of this system is that p_c is not a monotonic function of $\langle K \rangle$, having a maximum value of $(1-\rho)/e$ at $\langle K \rangle = e/(1-\rho)$ and approaching zero as $\langle K \rangle$ approaches infinity. Thus if p is held fixed at any value between zero and $(1-\rho)/e$, the system will undergo two transitions as $\langle K \rangle$ is increased from zero. The system will begin in the EP regime (i.e. $p > p_c$), undergo a transition to subcritical behavior at some $\langle K \rangle$, then reenter
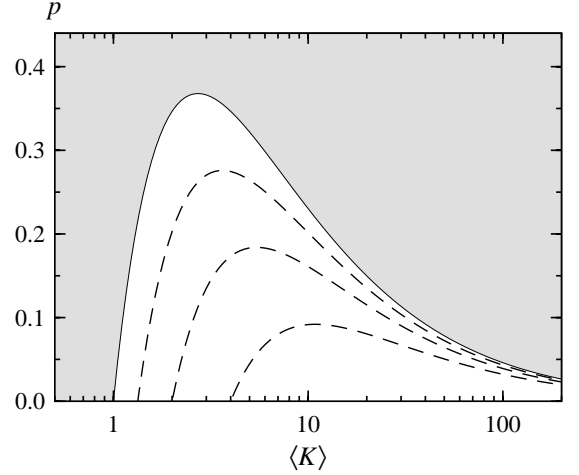

FIG. 3: Phase diagram for EP on random digraphs, where damage spreads along each directed link with probability p and a node is guaranteed to get damaged in the special case that all of its inputs are connected to damaged nodes. All nodes with zero inputs are initially damaged, and the other nodes are initially damaged with probability ρ. The gray area bounded by a solid line shows the region where EP occurs for $\rho = 0$ and the dashed lines show the EP transition when ρ has the values $1/4$, $1/2$, and $3/4$, respectively.

the EP regime for a higher value of $\langle K \rangle$. The calculated phase diagram is shown in Fig. 3 and has been verified by direct numerical simulations of avalanches. Roughly speaking, at low $\langle K \rangle$ EP occurs due to the high density of initially damaged nodes with no inputs. At high $\langle K \rangle$, on the other hand, EP occurs due to the high probability of nodes being damaged because of their large number of inputs.

B. The probability of complete coverage

An important quantity associated with EP is the probability of an avalanche yielding complete coverage of the system; i.e., the probability that all sites are damaged by the UBA so that $u = 0$. Let $P_{\text{cc}}(N, q; n_0, n_{0*})$ denote the probability that a UBA on a random network will yield complete coverage for a system with a given network size N, a given damage control function q, and starting with particular values of n_0 and n_{0*}. For future convenience we also define $P_{\text{cc}}(N, q)$ to be the probability for complete coverage assuming that each node is initially damaged with probability $1 - q(1)$ and we average over the corresponding probability distribution for n_{0*}.

To calculate $P_{\text{cc}}(N, q; n_0, n_{0*})$, we note that

$$P_{\text{cc}}(N, q; m, 0) = 0 \quad \text{if } m > 0, \quad (33)$$

since the process stops when $n_{0*} = 0$. We also have

$$P_{\text{cc}}(N, q; 0, m) = 1 \quad \text{for any } m, \quad (34)$$

since updating can never create 0s. These values of $P_{\rm CC}$ can be used for recursive calculation of $P_{\rm CC}$. Let $n_{0,0*}$ denote $n_0 + n_{0*}$, or $Nx_{0,0*}$. Performing steps 3–5 (from Section II A) one time decreases $n_{0,0*}$ by 1 as described by Eqs. (5)–(8). This means that $P_{\rm CC}(N, q; n_0, n_{0*})$ can be calculated for all $n_{0,0*} = m$ if $P_{\rm CC}(N, q; n_0, n_{0*})$ is known for all $n_{0,0*} = m - 1$. The recursion starts at $n_{0,0*} = 0$ with $P_{\rm CC}(N, q; 0, 0) = 1$ and uses the boundary conditions $P_{\rm CC}(N, q; n_{0,0*}, 0) = 0$ and $P_{\rm CC}(N, q; 0, n_{0,0*}) = 1$ for $n_{0,0*} > 0$.

For large N, $P_{\rm CC}$ can be calculated in the framework of a continuous approximation. Let $p(n_{0,0*}, c)$ denote a continuous version of $P_{\rm CC}(N, q; n_0, n_{0*})$. Then, the boundary conditions $P_{\rm CC}(N, q; n_{0,0*}, 0) = 0$ and $P_{\rm CC}(N, q; 0, n_{0,0*}) = 1$ are expressed as

$$p[n_{0,0*}, c_{\max}(x_{0,0*})] = 0, \qquad (35)$$

and

$$p(n_{0,0*}, 0) = 1, \qquad (36)$$

where

$$c_{\max}(x_{0,0*}) = \frac{n_{0,0*}}{q(x_{0,0*})}. \qquad (37)$$

In the continuous approximation, the recurrence relation that can be derived from Eqs. (5)–(8) is transformed to a partial differential equation. In such an update, the change $n_{0,0*}$ decreases by unity and, for large N, the change in c is much less than c itself. In the continuous approximation, this means that $p(n_{0,0*}, c)$ satisfies a partial differential equation of the form

$$\frac{\partial p}{\partial n_{0,0*}} = h_1(n_{0,0*}, c)\frac{\partial p}{\partial c} + h_2(n_{0,0*}, c)\frac{\partial^2 p}{\partial c^2}, \qquad (38)$$

where $h_1(n_{0,0*}, c)$ and $h_2(n_{0,0*}, c)$ are functions to be determined. This is recognizable as a 1D diffusion equation in which $n_{0,0*}$ plays the role of time and c the role of space. Note that later times in the diffusion equation correspond to earlier stages of the UBA, since $n_{0,0*}$ decreases as nodes are converted to 1s. The boundary conditions on the diffusion are given by Eqs. (35) and (37). We are interested in computing $p(n_{0,0*}, c)$ for values of $n_{0,0*}$ and c corresponding to $n_{0*} = n_{0*}^{\rm i}$ and $n_1 = 0$.

The fact that the average of c is constant means that the coefficient of the drift term in the diffusion equation must vanish; i.e., $h_1(n_{0,0*}, c) = 0$. The diffusion coefficient, $h_2(n_{0,0*}, c)$, is given by

$$h_2 = \tfrac{1}{2}\sigma^2(c'), \qquad (39)$$

where $\sigma^2(c')$ is the variance of c' when a fixed c is updated.

Using Eqs. (16) and (39) and converting g's to q's, we find

$$\frac{\partial p}{\partial n_{0,0*}} = \frac{c}{2N[q(x_{0,0*})]^2}\frac{dq(x)}{dx}\bigg|_{x=x_{0,0*}}\frac{\partial^2 p}{\partial c^2}. \qquad (40)$$

The large N behavior of Eq. (40), with the boundary conditions in Eqs. (35) and (37), can be found by expanding $q(x)$ around $x = 0$. If $q(x)$ is well-behaved, such an expansion can be written as

$$q(x) = \alpha_1 x - \alpha_2 x^2 + \mathcal{O}(x^3). \qquad (41)$$

This expansion can always be performed if the probability $P(K)$ for a node to have K inputs decays as least as fast as K^{-4} and in the case that p_K decays slower than K^{-4} but faster than K^{-3}, only the residue term can be affected. See Appendix A. In particular, the expansion is always valid if K has a maximal value.

The most interesting case in terms of asymptotic behavior is when α_1 is close to 1 and α_2 is positive. With suitable N-dependent transformations of p and its arguments, described in Appendix B, the large N behavior of Eq. (40) can be expressed in terms of a function $\tilde{p}(\tilde{t}, \tilde{y})$ determined by the differential equation

$$\frac{\partial \tilde{p}}{\partial \tilde{t}} = \frac{1}{2}\frac{\partial^2 \tilde{p}}{\partial \tilde{y}^2}, \qquad (42)$$

with the boundary conditions

$$\tilde{p}(\tilde{t}, 1/\tilde{t}) = 0 \quad \text{for } \tilde{t} < 0 \qquad (43)$$

and

$$\lim_{\tilde{t} \to -\infty} \tilde{p}(\tilde{t}, \tilde{y}) = \tilde{y} \quad \text{for } \tilde{y} \geq 0. \qquad (44)$$

The Crank–Nicholson method can be used to calculate $\tilde{p}(\tilde{t}, \tilde{y})$ numerically in an efficient way. (See, e.g., [35].)

Appendix B shows that the probability for complete coverage is given by

$$P_{\rm CC}(N, q) \approx \tilde{N}^{-1/3}\tilde{p}[0, \tilde{N}^{1/3}(1 - \alpha_1)], \qquad (45)$$

where $\tilde{N} = \alpha_1 N/\alpha_2$. The calculation assumes that the avalanche is initiated on the nodes whose outputs are independent of their inputs, as accounted for in $q(1)$.

To our knowledge, the critical point for EP has not been investigated previously in its own right. Two special cases have been studied, however. First, results for numbers of frozen and unfrozen nodes in critical RBNs can be mapped to an EP process, as discussed in Section IV. In this context, frozen nodes in the network are considered to be the damaged nodes of the UBA, and the scaling with N of the number of unfrozen nodes at the phase transition has been investigated for certain class of RBNs [25, 29].

Second, in the special case that $q(x) = x$, the exact result

$$P_{\rm CC}(N, x \mapsto x; n_0, n_{0*}) = \frac{n_{0*}}{n_0 + n_{0*}} \qquad (46)$$

is obtained. [See Eq. (B28) in Appendix B.] This means that the probability for complete coverage is exactly $n_{0*}^{\rm i}/N$. The simplest realization of $q(x) = x$ is provided

by a network of one-input nodes with rules that copy the input state. Such networks have strong connections to random maps from a set of N elements into itself. A map T is derived from a network of one-input nodes by letting each node map to the node from where its input is taken. In this picture, the damage originating from one initially damaged node i, corresponds to the set of nodes j such that $T^k(j) = i$ for some $k \geq 0$ (where T^k denotes the kth iterate of T). Such a j is called a *predecessor* to i. See, e.g., Ref. [34] for an overview of the theory of random maps and see Refs. [36, 37] for results on predecessors in random maps. See Appendix E for analytic results that relate UBA to random maps.

C. On the number of damaged nodes in random networks

In the Sections III A and III B we focused on determining the parameters that lead to EP (a vanishing fraction of undamaged nodes large N limit) and on the probability that the number of undamaged nodes will be exactly zero (complete coverage). We now consider the full probability distribution for the number of nodes damaged in an avalanche in a manner that provides a suitable base for understanding both SP and EP in random networks. The calculational strategy involves considering a given set of n nodes to be the damaged set and computing the probability that this is both consistent with all of the Boolean rules and the probability that the avalanche will actually cover the whole set. The probability of consistency is calculated via elementary combinatorics. The probability of reaching the whole set is precisely the probability of complete coverage for an avalanche on the sub-network of n candidate nodes. For this we can directly apply the results of the last section. For the purposes of explaining the calculation, we refer to the selected set of n nodes as the *candidate set*.

We let $P_{n,N}(n)$ denote the probability that n nodes will be damaged in an avalanche, averaged over the ensemble of N-node networks with a rule distribution characterized by a given damage propagation function g or the corresponding damage control function q. We assume that the avalanche is initiated by randomly selecting ℓ nodes to set to 0*, regardless of their Boolean rules, then setting to 0* all nodes with rules that always output 1 for any inputs. The set of ℓ initially damaged nodes must be a subset of the candidate set. The probability that the candidate set contains all of the nodes with "always 1" rules will be taken into account by the value of $g(0)$ in the expression below for the consistency probability. We use the notation $\binom{m}{k}$ for the usual binomial coefficient (the number of combinations of k objects chosen from a set of m objects).

The probability $P_{n,N}(n)$ can be expressed as

$$P_{n,N}(n) = \binom{N-\ell}{n-\ell} P_c(n, \ell; N) \, P_{CC}^1(n, \ell; N), \quad (47)$$

where $P_c(n, \ell; N)$ and $P_{CC}^1(n, \ell; N)$ are defined below. The binomial factor counts the number of different sets of $n - \ell$ nodes that could be damaged in a process corresponding to a given set of ℓ nodes that are initially damaged without regard to their rules. $P_c(n, \ell; N)$ is the probability that a given choice of $n - \ell$ nodes assumed to be damaged by the avalanche will constitute a final state that is consistent with the Boolean rules for each node, including the nodes that are initially damaged because their rules require it. $P_{CC}^1(n, \ell; N)$ is the probability that the avalanche will not die out before damaging all n nodes. This factor is necessary to avoid counting final states that contain loops of damaged nodes consistent with the rules but unreachable because damage cannot spread to the loop from any nodes outside the loop.

Consistency with the Boolean rules requires that the given set of $n - \ell$ nodes damaged in the avalanche have inputs that cause them to be damaged. In a random network, the probability that any single node will be damaged is $g(x_1)$, where x_1 is the fraction of damaged nodes. Similarly, the probability that any node will *not* be damaged is $1 - g(x_1)$. We are considering candidate sets of damaged nodes with $x_1 = n/N$. Thus we have

$$P_c(n, \ell; N) = [g(n/N)]^{n-\ell}[1 - g(n/N)]^{N-n}. \quad (48)$$

The computation of $P_{CC}^1(n, \ell; N)$ involves the rule distribution on the restricted network formed by the candidate set with all inputs from the undamaged nodes removed. This distribution, $g^1(x)$, is different from $g(x)$ because P_c already accounts for rules that are not consistent with the pattern of damage. Thus the spreading of damage on the n-node network involves $g(nx/N)$, the probability that a rule outputs 1 when a fraction x of the n-node candidate set is damaged. The probability must be normalized such that it goes to unity when x goes to 1. (We know that a node in the n-node set should get damaged if all of its inputs are damaged.) Thus we have

$$g_{N,n}^1(x) = \frac{g(nx/N)}{g(n/N)} \quad (49)$$

or, equivalently,

$$q_{N,u}^1(x) = \frac{q[u/N + (1 - u/N)x] - q(u/N)}{1 - q(u/N)}. \quad (50)$$

(Recall that $u = N - n$ is the number of undamaged nodes after an avalanche.)

There are two cases of interest for the probability of complete coverage of the candidate set. For EP, $g(0) > 0$ and the fixed number ℓ of nodes arbitrarily selected for damage is irrelevant compared to the finite fraction of nodes with rules that produce damage for any combination of inputs. In this case, we assume $\ell = 0$, which allows reduction of P_{CC} to its two-argument form defined at the beginning of Section III B:

$$P_{CC}^1(n, 0; N) = P_{CC}(n, q_{N,N-n}^1). \quad (51)$$

For SP, we have $g(0) = 0$ so the avalanche must be initiated with a nonzero value of ℓ. In this case we have

$$P_{cc}^1(n, \ell; N) = P_{cc}(n, q_{N,N-n}^1; n - \ell, \ell). \qquad (52)$$

Note that $P_{cc}^1(n, \ell; N)$ depends on N only through q^1.

For notational convenience, we now let P_{cc}^1 stand for whichever expression on the right-hand side of Eqs. (51) or (52) is relevant, and we use u where $N - n$ would be the strictly proper form. By combining Eqs. (47) and (48), we get

$$P_{n,N}(n) = \binom{N - \ell}{n - \ell}[g(n/N)]^{n-\ell}[1 - g(n/N)]^u P_{cc}^1. \qquad (53)$$

To make some important features of Eq. (53) apparent, we introduce the functions

$$\rho(n) = \frac{n^n}{e^n n!}, \qquad (54)$$

$$\tau(n, k) = \frac{n!}{n^k(n-k)!}, \qquad (55)$$

and

$$G(x) = \left(\frac{g(x)}{x}\right)^x \left(\frac{1 - g(x)}{1 - x}\right)^{1-x}. \qquad (56)$$

Then Eq. (53) can be rewritten as

$$P_{n,N}(n) = \frac{\rho(n)\rho(u)}{\rho(N)} \frac{\tau(n, \ell)}{\tau(N, \ell)} \left(\frac{n/N}{g(n/N)}\right)^\ell$$
$$\times [G(n/N)]^N P_{cc}^1. \qquad (57)$$

Stirling's formula,

$$n! \approx \sqrt{2\pi n}\, \frac{n^n}{e^n}, \qquad (58)$$

yields

$$\rho(n) \approx \frac{1}{\sqrt{2\pi n}} \qquad (59)$$

and

$$\frac{\rho(n)\rho(u)}{\rho(N)} \approx \frac{1}{\sqrt{2\pi n u/N}}. \qquad (60)$$

The factor $\tau(n, \ell)/\tau(N, \ell)$ is approximately 1 for large n and satisfies

$$\frac{\tau(n, \ell)}{\tau(N, \ell)} \leq 1 \qquad (61)$$

for $n \leq N$, with equality if $n = N$ or $\ell = 1$ or $\ell = 0$. The only factors in Eq. (57) that can show exponential dependence on N are the G and P_{cc}^1 factors. Because P_{cc}^1 is a probability (and therefore cannot exceed unity)

and $G(x) \leq 1$ with equality if and only if $g(x) = x$, $P_{n,N}(n)$ vanishes exponentially as N goes to infinity for any fixed n/N such that $g(n/N) \neq n/N$. This is consistent with the above result that the probability of an avalanche stopping with $x_1 \neq g(x_1)$ is vanishingly small. [See Eqs. (19) and (25).]

For EP, we are interested in the number of undamaged nodes, u. We let

$$P_{u,N}(u) = P_{n,N}(N - u) \qquad (62)$$

and

$$Q(x) = G(1 - x)$$
$$= \left(\frac{1 - q(x)}{1 - x}\right)^{1-x} \left(\frac{q(x)}{x}\right)^x. \qquad (63)$$

For EP, $g(0) > 0$ and a fixed ℓ is irrelevant when $N \to \infty$. Hence, we let $\ell = 0$ and rewrite Eq. (57) as

$$P_{u,N}(u) = \frac{\rho(n)\rho(u)}{\rho(N)}[Q(u/N)]^N P_{cc}^1. \qquad (64)$$

To some respects, $P_{u,N}$ is similar to $P_{n,N}$: the factor $[\rho(u)\rho(n)]/\rho(N)$ is fully symmetric with respect to interchange of n and u; and the role of $G(n/N)$ in Eq. (57) is identical to the role of $Q(u/N)$ in Eq. (64). However, the behavior of P_{cc}^1 for $n \ll N$ given by Eq. (52) is significantly different from the behavior of P_{cc}^1 for $u \ll N$ given by Eq. (51).

For EP, we consider damage control functions $q(x)$ that can be expanded according to Eq. (41). For supercritical EP, with $\alpha_1 < 1$, $P_{u,N}(u)$ decays exponentially with u. In Appendix D 1, we demonstrate that

$$\lim_{N \to \infty} P_{u,N}(u) = (1 - \alpha_1)\frac{(u\alpha_1)^u}{u!}e^{-u\alpha_1} \qquad (65)$$

$$\approx \frac{1 - \alpha_1}{\sqrt{2\pi}}e^{u(1-\alpha_1)}\alpha_1^u u^{-1/2}. \qquad (66)$$

For critical EP, Eq. (45) gives

$$P_{cc}^1(n, 0; N) = P_{cc}(n, q_{N,N-n}^1)$$
$$\approx \tilde{n}^{-1/3}\tilde{p}[0, \tilde{n}^{1/3}(1 - \alpha_1^1)], \qquad (67)$$

where $\tilde{n} \equiv \alpha_1^1 n/\alpha_2^1$ and α_1^1 and α_2^1 are the first two coefficients of the power series expansion of $q^1(x)$ about $x = 0$. With $\alpha_1 = 1$ and $\alpha_2 > 0$, a Taylor expansion of $\log Q(x)$ about $x = 0$ gives

$$Q(x) \approx \exp\left(-\frac{\alpha_2^2 x^3}{2}\right) \qquad (68)$$

for small x. This yields that the typical number of undamaged nodes, u, scales like $N^{2/3}$. In Appendix D 2, we derive the asymptotic distribution of u for large N. With $\tilde{u} = \tilde{N}^{-2/3}u = (\alpha_2/N)^{2/3}u$, we find that the large N limit of the probability density for $\tilde{u}$ is

$$P(\tilde{u}) = \frac{\exp(-\frac{1}{2}\tilde{u}^3)}{\sqrt{2\pi\tilde{u}}}\tilde{p}(0, 2\tilde{u}). \qquad (69)$$

Eq. (57) is suitable for understanding SP as well as EP. For SP, $g(0) = 0$ and $\ell > 0$. In the large N limit, SP is a branching process with a Poisson distribution in the number of branches from each node. The average number of branches per node is given by the derivative of $g(x)$ at $x = 0$, because $\lim_{x \to 0} g(x)/x$ is the average number of nodes that will be damaged in one update as a direct consequence of damaging a single node in the large network limit. In Appendix C, we re-derive known results on SP in the framework of our formalism.

IV. AN APPLICATION: FROZEN NODES IN RANDOM BOOLEAN NETWORKS

An important application of our results on EP in random networks is the determination of the size distribution for the set of unfrozen nodes in 2-input random Boolean networks, a subject of interest since the introduction of the Kauffman model in 1969 [17]. The Kauffman model was originally proposed as a vehicle for studying aspects of the complex dynamics of transcriptional networks within cells.

In a Boolean network, there are usually some nodes that will reach a fixed final state after a transient time regardless of the initial state of the network. For most random Boolean networks, nearly all of these nodes can be found by a procedure introduced in Ref. [26] and applied numerically in Ref. [27]. We refer to nodes identified by this procedure as *frozen*.

The nodes that cannot be identified as frozen are labeled *unfrozen*. Their output may switch on and off for all time or simply have different values on different attractors of the network dynamics. A frozen node will always reach its fixed final state regardless of the initial state of the network. The converse is not true: an unfrozen node can have a fixed final state that is independent of the initial state due to correlations that are not accounted for in the identification procedure for frozen nodes. In a typical random Boolean network, the number of nodes that are mislabeled in this sense is negligible [27]. For the purposes of investigating dynamics of the network at long times, one is interested in the size of the unfrozen set.

The procedure for identification of the frozen nodes starts by marking all nodes with a constant output function as frozen. There may then be nodes that, as a consequence of receiving one or many inputs from frozen nodes, will also produce a constant output. These nodes are also marked as frozen, and the process continues iteratively until there are no further nodes that can be identified as frozen.

We note here that the process of finding frozen nodes in a RBN can often be framed as a UBA, where the property of being frozen corresponds to damage. That is, the process of identifying frozen nodes involves continually checking all nodes to see whether their inputs are frozen in such a way that they themselves become frozen, a pro-cess which satisfies the conditions for UBA. The damage propagation and damage control functions for the UBA are determined by the relative weights of different Boolean logic functions in the RBN. By changing these weights, one can observe a transition in the dynamical behavior of RBNs corresponding precisely to the EP transition in the UBA. We consider here RBNs with exactly two inputs at each node, with some explicit choices for the weights of the Boolean logic functions that permit observation of both sides of the transition.

The only restriction required for mapping the freezing of nodes in a RBN to a UBA system is that the logic functions in the RBN be symmetric with respect to the probability of freezing being due to TRUE and FALSE inputs. That is, the probability that a node with a certain set of frozen inputs will itself be frozen should not depend on the values of the frozen inputs. This condition is satisfied for the most commonly investigated classes of rule distributions, where there is a given probability p for obtaining a 1 at each entry in the truth table for each rule. If the above mentioned symmetry condition were violated, it would be necessary to distinguish nodes frozen TRUE from nodes frozen FALSE, which would mean that the state of a node could not be specified by a binary variable. For the rest of this section we consider only RBNs that respect the symmetry condition.

It is useful to distinguish different types of Boolean logic functions. A *canalizing* rule is one for which the output is independent of one of the inputs for at least one value of the other input. Among the 16 possible 2-input Boolean rules, 2 rules are constant ("always on" or "always off"), 12 rules are non-constant and canalizing, and 2 rules are non-canalizing (XOR and not-XOR). The original version of the Kauffman model assumes that all 2-input Boolean rules are equally likely, which turns out to give critical dynamics.

Let p_i denote the probability that a randomly selected node's output is frozen if exactly i of its inputs are frozen. The damage propagation function $g(x)$ can be expressed directly in terms of p_i:

$$g(x) = p_0(1-x)^2 + 2p_1 x(1-x) + p_2 x^2. \quad (70)$$

Nodes with constant rules are guaranteed to be frozen. (These nodes will initiate the UBA.) Nodes with non-constant canalizing rules are unfrozen if both inputs are unfrozen, and they are frozen with probability $1/2$ if exactly one randomly selected input is frozen. Nodes with rules that are non-canalizing become frozen if and only if both of their inputs are frozen. Finally, if both inputs are frozen, the output of any 2-input rule is frozen. Thus for the 2-input Kauffman model, $p_0 = 1/8$, $p_1 = 1/2$, and $p_2 = 1$.

If the two non-canalizing rules in the 2-input Kauffman model are replaced by canalizing rules, p_1 becomes $9/16$, whereas p_0 and p_2 are unchanged. Such networks exhibit supercritical EP. To get a subcritical network, we replace two of the canalizing rules with non-canalizing rules and get $p_1 = 7/16$. (Note that some care must be taken to
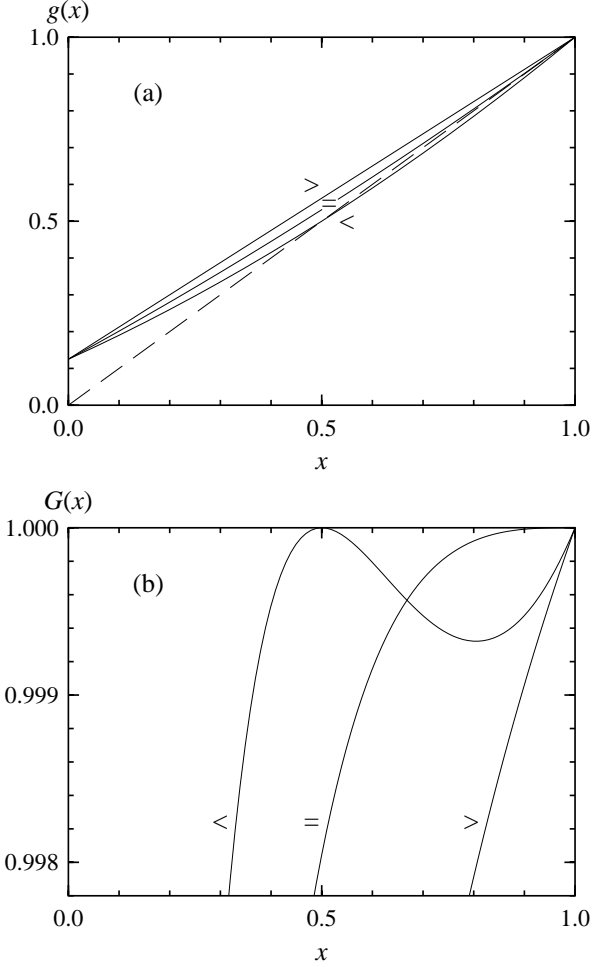
FIG. 4: The functions (a) $g(x) \equiv 1 - q(1 - x)$ and (b) $G(x) \equiv Q(1 - x)$ for three 2-inputs rule distributions. All three distributions have $p_0 = 1/8$ and $p_2 = 1$, whereas p_1 takes the values 7/16, 1/2, and 9/16. The case that has $p_1 = 1/2$ (marked with =) is critical with respect to EP and corresponds to the propagation of frozen node values in the original Kauffman model. The other cases $p_1 = 7/16$ (<) and $p_1 = 9/16$ (>) are subcritical and supercritical, respectively. The dashed line in (a) shows the identity function $x \mapsto x$.

FIG. 5: The probability density distribution $NP_{n,N}(n)$ with respect to the fraction of nodes (n/N) involved in an avalanche. The rule distributions have the same $g(x)$ as displayed in Fig. 4, showing rule distributions that are (<) subcritical, (=) critical, and (>) supercritical with respect to EP. The displayed networks sizes, N, are 10 (large dots), 100 (small dots), 10^3 (bold line), 10^4, 10^5, and 10^6 (gradually thinner lines).

maintain the TRUE–FALSE symmetry mentioned above.) The functions $g(x)$ and $G(x)$ for critical, supercritical, and subcritical rule distributions are shown in Fig. 4.

As can be seen from Fig. 4, a small change in $g(x)$ may lead to a qualitative change in $G(x)$ for rule distributions close to criticality. Such changes have a strong impact on the avalanche size distribution for large N. Figure 5 shows the probability density distribution of the fraction, n/N, of nodes that are affected by avalanches in networks with the above mentioned rule distributions. The probability distributions are obtained by recursive calculation of the distribution of n_{0*} as n_1 increases. The recurrence relations are obtained from Eqs. (5)–(8) and the result is exact up to truncation errors. To verify these calcula-
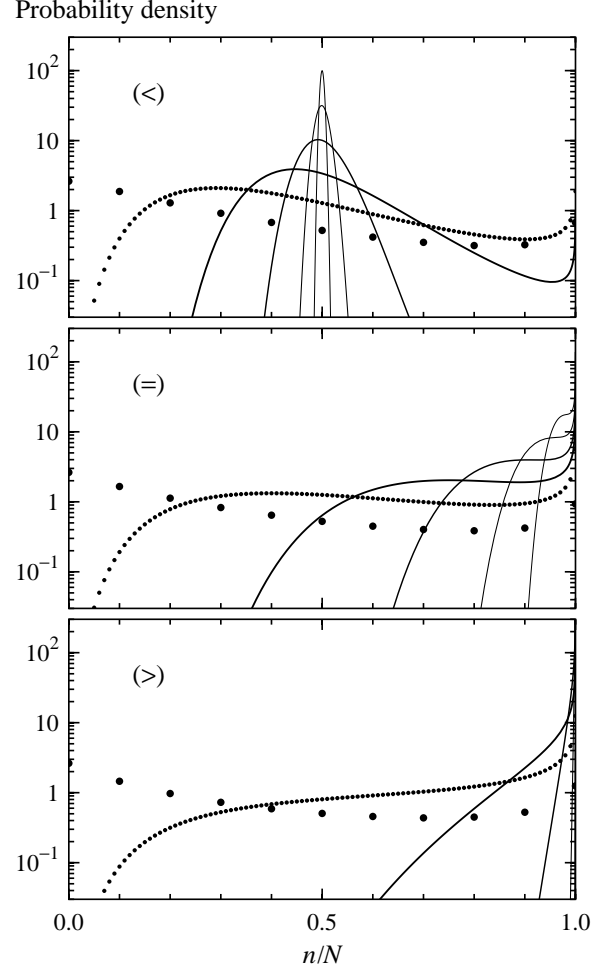
tions, we generated 10^6 random Boolean networks of size $N = 10^3$ for each of the above described rule distributions. The distributions in the numbers of frozen nodes in those networks are displayed in Fig. 6.

In Fig. 7, the probability distributions of the number of undamaged nodes, u, are shown in comparison to the asymptotic results in Eqs. (65) and (69). Our analytic results are strengthened by the data in Fig. 7 as the distributions for finite networks approaches the predicted asymptotes. Finite size effects are clearly visible in the critical case even for network sizes as big as $N = 10^6$, whereas convergence in the supercritical case is achieved for $N \gtrsim 10^3$.

Kaufman, Mihaljev, and Drossel studied distributions of unfrozen nodes in 2-input critical RBNs using a method

Probability density
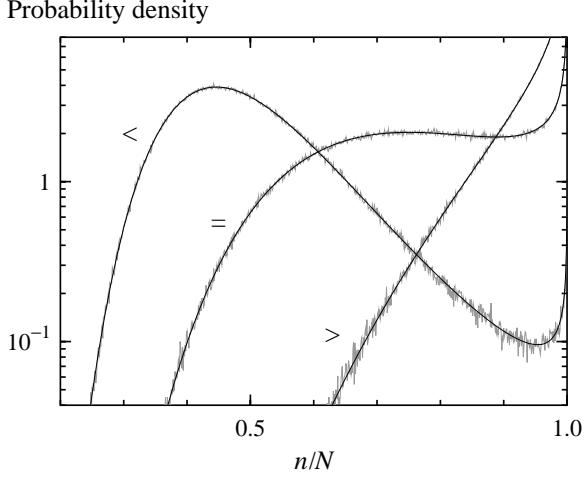

Probability density

Probability

FIG. 6: A numeric comparison between analytic calculations (black lines) and explicit reductions of random Boolean networks (gray lines). For both cases, the probability density distribution $NP_{n,N}(n)$ is displayed as a function of n/N. The rule distributions have the same $g(x)$ as displayed in Figs. 4 and 5, showing rule distributions that are ($<$) subcritical, ($=$) critical, and ($>$) supercritical with respect to EP. The UBA rule distributions are realized in random Boolean networks by rule distributions with the following respective selection probabilities: $1/8$, $1/4$, $5/8 - p_r$, and p_r for a constant rule, a rule that depends on exactly 1 input, a canalizing rule that depends on 2 inputs, a 2-input reversible rule. The values of p_r are ($<$) 0, ($=$) 1/8, and ($>$) 1/4. For each rule distribution, 10^6 networks were tested.

similar to ours in that differential equations for populations of different types of nodes are developed from a discrete process in which frozen nodes are identified by the propagation of information from their inputs [29]. Their result for the numbers of unfrozen nodes in 2-input critical RBNs corresponds to a particular application of Eq. (69). In Ref. [29], the function corresponding to $P(\tilde{u})$ [which they call $G(y)$] is determined by running a stochastic process and a numerically motivated approximation is proposed:

$$P(\tilde{u}) \approx 0.25 \exp(-\tfrac{1}{2}\tilde{u}^3) \frac{1 - 0.5\sqrt{\tilde{u}} + 3\tilde{u}}{\sqrt{\tilde{u}}}. \quad (71)$$

The scaling law $P(\tilde{u}) \propto \tilde{u}^{-1/2}$ for small $\tilde{u}$ is also derived analytically in Ref. [29].

For large x, Eqs. (42)–(44) imply $\tilde{p}(0,x) \propto x$ for large positive x. This means that

$$P(\tilde{u}) \approx \sqrt{\frac{2\tilde{u}}{\pi}} \exp(-\tfrac{1}{2}\tilde{u}^3) \quad (72)$$

for large $\tilde{u}$. Thus the large $\tilde{u}$ limit of Eq. (71) differs from the exact result by a factor of $(3/4)\sqrt{\pi/2}$, an underestimate of about 6%.

We are able to improve further on Eq. (71) by numerical investigations of $\tilde{p}(0,x)$ calculated by the Crank–
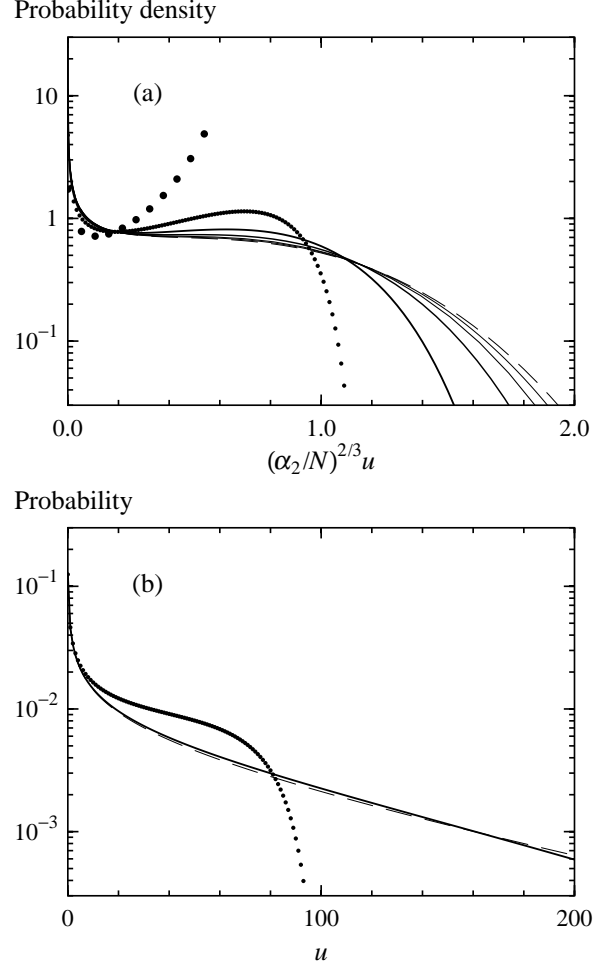
FIG. 7: Rescaled versions of the probability distributions displayed in Fig. 5: (a) the probability density for the critical case, with respect to the rescaled number of undamaged nodes, $\tilde{u} \equiv (\alpha_2/N)^{2/3} u = u/(4N^{2/3})$; (b) the probability distribution $P_{u,N}(u)$ for the supercritical case without rescaling. The displayed networks sizes, N, are 10 (large dots), 100 (small dots), 10^3 (bold line), 10^4, 10^5, and 10^6 (gradually thinner lines). The analytically derived asymptotes are shown as dashed lines. In (b), the distributions for networks of sizes 10^4, 10^5, and 10^6 are not plotted because they are indistinguishable from the asymptotic curve.

Nicholson method (see, e.g., [35]) using Eqs. (42)–(44). We find that the high-precision numerical results are fit by the function

$$P(\tilde{u}) \approx \sqrt{\frac{2\tilde{u}}{\pi}} \exp(-\tfrac{1}{2}\tilde{u}^3)\left(1 + \frac{1}{3.248\tilde{u} + 4.27\tilde{u}^2 + 4.76\tilde{u}^3}\right) \quad (73)$$

with a relative error that is maximally 0.25% and vanishing for large $\tilde{u}$.

By explicitly keeping track of the populations of nodes with each of the different types of Boolean logic functions as links from frozen nodes are deleted, Kaufman, Mihal-

jev, and Drossel [29] also derive results for other quantities, such as the number of links in the sub-network of unfrozen nodes. The EP formalism described above can be applied once again to investigate these additional quantities in a broader class of networks. Detailed results for RBNs with various degree distributions will be presented elsewhere.

V. SUMMARY AND DISCUSSION

Unordered binary avalanches can in some cases lead to damage on every node or almost every node of a network, a phenomenon we have dubbed *exhaustive percolation*. We have studied a broad class of random networks that can exhibit EP. We have shown how to calculate the probability $P_{\mathrm{cc}}(N)$ that complete coverage occurs (i.e that all nodes are damaged) and also derived expressions for the probability distribution $P(u)$ of the number of undamaged nodes, u, in the large N limit when EP does occur. A logical curiosity in our approach is the fact that the calculation of $P(u)$ involves application of the P_{cc} result to subnetworks containing candidate sets of damaged nodes.

Our primary results flow from the realization that all of the relevant information about a UBA defined on a random network is contained in the damage propagation function $g(x)$ or, equivalently, the damage control function $q(x)$. We derive scaling law exponents and exact results for the distribution of u that are valid for a broad class of random networks and Boolean rule distributions in the EP regime and for networks at the EP critical point. This class includes the UBAs that determine the set of frozen nodes in RBNs with more than two inputs per node and therefore constitute a generalization of the results on the set of unfrozen nodes in RBNs presented in Ref. [29]. Interestingly, the asymptotic behavior found in Ref. [29]

for the distribution of u at the critical point is shown to be valid for a broad class of network problems.

For networks outside the above mentioned class but within the framework of UBA, we find connections to previous work on Galton–Watson processes [32] and random maps [36]. The central result of our investigations is displayed in Eqs. (66) and (69), which provide explicit formulas for the probability of finding u undamaged nodes after an avalanche runs to completion. The out-degree distributions of the networks described by our formulas are all Poissonian, but the in-degree distributions may have different forms, including power laws, so long as the probability of having in-degree K decays faster than K^{-3}. The exact nature of the EP transition on networks with broader in-degree distributions is an interesting issue for future research. Further work is also needed to handle correlations between input links to different nodes, a situation that arises, for example, in random regular graphs or networks with scale free out-degree distributions.

Our original motivation for studying EP arose from attempts to understand the dynamical behavior of RBNs. We have described one nontrivial example of how the EP formalism is relevant: the calculation of the probability distribution for the number of unfrozen nodes in any RBN with a rule distribution that leads to a given damage control function q for the associated UBA. The problem of determining how many of the unfrozen nodes are actually relevant for determining the attractor structure of the RBN can also be framed as an EP problem, which will be addressed in a separate publication.

[1] R. Kinney, P. Crucitti, R. Albert, and V. Latora, Eur. Phys. J. B **46**, 101 (2005).
[2] M. L. Sachtjen, B. A. Carreras, and V. E. Lynch, Phys. Rev. E **61**, 4877 (2000).
[3] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).
[4] R. Cohen, S. Havlin, and D. ben-Avraham, Phys. Rev. Lett. **91**, 247901 (2003).
[5] H. W. Hethcote, SIAM Rev. **42**, 599 (2000).
[6] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[7] A. L. Lloyd and R. M. May, Science **292**, 1316 (2001).
[8] T. R. Hughes, M. J. Marton, A. R. Jones, C. J. Roberts, R. Stoughton, C. D. Armour, H. A. Bennett, E. Coffey, H. Dai, Y. D. He, et al., Cell **102**, 109 (2000).
[9] P. Rämö, J. Kesseli, and O. Yli-Harja (2006), to appear in J. Theor. Biol., doi:10.1016/j.jtbi.2006.02.011.
[10] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[11] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).
[12] Y. Moreno, J. B. Gómez, and A. F. Pacheco, Europhys. Lett. **58**, 630 (2002).
[13] D. J. Watts, Proc. Natl. Acad. Sci. USA **99**, 5766 (2002).
[14] A. E. Motter, Phys. Rev. Lett. **93**, 098701 (2004).
[15] P. Crucitti, V. Latora, and M. Marchiori, Phys. Rev. E **69**, 045104(R) (2004).
[16] D. Watts, *Small worlds: the dynamics of networks between order and randomness* (Princeton University Press, 1999).
[17] S. A. Kauffman, J. Theor. Biol. **22**, 437 (1969).
[18] B. Derrida and Y. Pomeau, Europhys. Lett. **1**, 45 (1986).
[19] U. Bastolla and G. Parisi, J. Theor. Biol. **187**, 117 (1997).
[20] M. Aldana-Gonzalez, S. Coppersmith, and L. P. Kadanoff, *Boolean Dynamics with Random Couplings*, in *Perspectives and Problems in Nonlinear Science*, edited by E. Kaplan, J. E. Marsden and K. R. Sreenivasan (Springer, 2003), p. 23.
[21] M. Aldana and P. Cluzel, Proc. Natl. Acad. Sci. USA **100**, 8710 (2003).

[22] B. Samuelsson and C. Troein, Phys. Rev. Lett. **90**, 098701 (2003).
[23] S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, Proc. Natl. Acad. Sci. USA **101**, 17102 (2004).
[24] B. Drossel, Phys. Rev. E **72**, 016110 (2005).
[25] J. E. S. Socolar and S. A. Kauffman, Phys. Rev. Lett. **90**, 068702 (2003).
[26] H. Flyvbjerg, J. Phys. A: Math. Gen. **21**, L955 (1988).
[27] S. Bilke and F. Sjunnesson, Phys. Rev. E **65**, 016129 (2002).
[28] U. Bastolla and G. Parisi, Physica D **115**, 203 (1998).
[29] V. Kaufman, T. Mihaljev, and B. Drossel, Phys. Rev. E **72**, 046124 (2005).
[30] M. Sahimi, *Heterogenous Materials I* (Springer, 2003).
[31] T. E. Harris, *The Theory of Branching Processes* (Springer Verlag, 1963).
[32] R. Otter, Ann. Math. Stat. **20**, 206 (1949).
[33] J. W. Essam, A. J. Guttmann, and K. De'Bell, J. Phys. A: Math. Gen. **21**, 3815 (1988).
[34] B. Bollabás, *Random Graphs* (Academic Press, London, 1985).
[35] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C, second ed.* (Cambridge University Press, 1992).
[36] B. Harris, Ann. Math. Statist. **31**, 1045 (1960).
[37] H. Rubin and R. Sitgreaves (1954), *Probability distributions related to random transformations on a finite set. Tech. Rep. 19A*, Applied Mathematics and Statistics Laboratory, Stanford University.
[38] M. Dwass, J. Appl. Prob. **6**, 682 (1969).

APPENDIX A: CALCULATION OF THE DAMAGE CONTROL FUNCTION

Let p_K denote the probability that a rule has K inputs, and let $P_0(K, m)$ denote the probability that the output value is zero of a rule with K inputs fed with m zeros and $K - m$ ones. Then, the the damage control function is

$$q(x) = \sum_{K=0}^{\infty} p_K \sum_{m=0}^{K} P_0(K, m) \binom{K}{m} x^m (1 - x)^{K-m}. \tag{A1}$$

Eq. (A1) can be written as

$$q(x) = a_0 + a_1 x + a_2 x^2 + \cdots \tag{A2}$$

where

$$a_i = \sum_{K=i}^{\infty} \sum_{m=0}^{i} p_K P_0(K, m)(-1)^{i-m} \binom{K}{m} \binom{K - m}{i - m}. \tag{A3}$$

The expansion in Eq. (A2) is well-defined up to the first term such that the sum in Eq. (A3) is not absolute convergent. The factor $\binom{K}{m}\binom{K-m}{i-m}$ scales like K^i for large K and $P_0(K, m) \leq 1$. Hence, a_i is well-defined if $\sum_{K=0}^{\infty} K^i p_K$ is convergent and this is true if p_K decays faster than K^{-i-1}.

In addition, the requirement that the output of each rule in the rule distribution is 1 if all of its inputs have the value 1, yields that $a_0 = 0$. Thus, the expansion

$$q(x) = \alpha_1 x - \alpha_2 x^2 + \mathcal{O}(x^3), \tag{A4}$$

is valid for all rule distributions such that p_K decays faster than K^{-4}. In the case that p_K decays slower than K^{-4} but faster than K^{-3}, only the residue term can be affected.

APPENDIX B: PROBABILITY FOR COMPLETE COVERAGE

Here, we assume that the expansion in Eq. (A4) is well-defined. Then, we get

$$\frac{\partial p}{\partial n_{0,0*}} = \frac{c}{2\alpha_1 N n_{0,0*}^2} \frac{\partial^2 p}{\partial c^2}[1 + \mathcal{O}(n_{0,0*}/N)], \tag{B1}$$

and

$$c_{\max}(x)/N = \frac{1}{\alpha_1} + \frac{\alpha_2}{\alpha_1^2} x + \mathcal{O}(x^2). \tag{B2}$$

To remove the dependence of $n_{0,0*}$ from the leading order term of the diffusion rate in Eq. (B1), we let $t = -1/n_{0,0*}$. By also letting $y = 1 - \alpha_1 c/N$, we rewrite Eq. (B1) to a form that easily can be rescaled as N grows. We get

$$\frac{\partial p}{\partial t} = \frac{1 - y}{2} \frac{\partial^2 p}{\partial y^2}[1 + \mathcal{O}(\tfrac{1}{Nt})]. \tag{B3}$$

and

$$y_{\min} = -\frac{\alpha_2}{\alpha_1 N t}[1 + \mathcal{O}(\tfrac{1}{Nt})], \tag{B4}$$

where $y = y_{\min}$ is the transformed value of $c_{\max}$. The boundary conditions are $p = 0$ for $y = y_{\min}$ and $p = 1$ for $y = 1$.

The N dependence of the leading order term of the boundary condition in Eq. (B4) can be removed by rescaling of y and t. Typically, $\alpha_2 > 0$ and this is the case that we will focus on. [Note that Eq. (25) means that α_2 must be nonnegative at the transition.] If $\alpha_2 = 0$, either $q(x) = x$ or $q(x) = x - \alpha_m x^m + \cdots$ with $m > 2$ (apart from some pathological special cases). The first case, $q(x) = x$, is a special case that is convenient for analytic calculation, whereas the latter case require calculations analogous to the calculations for $\alpha_2 > 0$. We will come back to the case $q(x) = x$.

For $\alpha_2 > 0$, we rescale y and t according to

$$\tilde{y} = \tilde{N}^{1/3} y \tag{B5}$$

and

$$\tilde{t} = \tilde{N}^{2/3} t, \tag{B6}$$

where

$$\tilde{N} = \frac{\alpha_1}{\alpha_2} N. \qquad (B7)$$

Then,

$$\frac{\partial p}{\partial \tilde{t}} = \frac{1}{2} \frac{\partial^2 p}{\partial \tilde{y}^2} \left(1 + \tilde{N}^{-1/3} \tilde{y}\right) \left[1 + \mathcal{O}\left(\frac{1}{\tilde{N}^{1/3} \tilde{t}}\right)\right], \qquad (B8)$$

where

$$\tilde{y}_{\min} = \tilde{t}^{-1} \left[1 + \mathcal{O}\left(\frac{1}{\tilde{N}^{1/3} \tilde{t}}\right)\right]. \qquad (B9)$$

The boundary conditions are $p = 0$ for $\tilde{y}_{\min}$ and $p = 1$ for $y = \tilde{N}^{1/3}$. The only plausible limit of p as $t \to -\infty$ is $p = y/\tilde{N}^{1/3}$. To get a motivation that is mathematically acceptable, one needs to relate the original integer based formulation of the problem in Eqs. (5)–(8). The large N behavior of Eq. (8),

$$\lim_{N \to \infty} U(\mathbf{s}, j) = 1/n_{0,0*}, \qquad (B10)$$

yields

$$\lim_{N \to \infty} P_{\mathrm{CC}}(N, q; n_0, n_{0*}) = \frac{n_{0*}}{n_{0,0*}}. \qquad (B11)$$

Eq. (B11) can be shown via induction. The induction is initiated by

$$\lim_{N \to \infty} P_{\mathrm{CC}}(N, q; 1, 0) = 0 \qquad (B12)$$

and

$$\lim_{N \to \infty} P_{\mathrm{CC}}(N, q; 0, 1) = 1, \qquad (B13)$$

which means that Eq. (B11) is true for $n_{0,0*} = 1$. To obtain the induction step, we assume that Eq. (B11) is true for $n'_{0,0*} = n_{0,0*} - 1$. Then, we get

$$\lim_{N \to \infty} P_{\mathrm{CC}}(N, q; n'_0, n'_{0*}) = \frac{n'_{0*}}{n_{0,0*} - 1} \qquad (B14)$$

which leads to

$$\lim_{N \to \infty} P_{\mathrm{CC}}(N, q; n_0, n_{0*}) = \frac{\langle n'_{0*} \rangle}{n_{0,0*} - 1} \qquad (B15)$$

$$= \frac{n_{0*} + n_0/n_{0,0*} - 1}{n_{0,0*} - 1} \qquad (B16)$$

$$= \frac{n_{0*}}{n_{0,0*}} \qquad (B17)$$

that completes the induction step. Eq. (B11) means that the value of p approaches a linear function of $\tilde{y}$ for $\tilde{t} = N^{2/3}/n_0$ as $N \to \infty$. Hence, the boundary condition for $t \to -\infty$ is $p = y/\tilde{N}^{1/3}$.

Rescaling of p according to

$$\tilde{p} = \tilde{N}^{1/3} p \qquad (B18)$$

gives the boundary condition

$$\lim_{t \to -\infty} \tilde{p} = \tilde{y}. \qquad (B19)$$

If Eq. (B19) is extended to be valid for all non-negative $\tilde{y}$, the boundary condition at $\tilde{y} = \tilde{N}^{1/3}$ can be dropped. In the limit of large N, Eq. (B8) becomes

$$\frac{\partial \tilde{p}}{\partial \tilde{t}} = \frac{1}{2} \frac{\partial^2 \tilde{p}}{\partial \tilde{y}^2}. \qquad (B20)$$

With $\tilde{p}$ is written on the form $\tilde{p}(\tilde{t}, \tilde{y})$, the boundary conditions are

$$\tilde{p}(\tilde{t}, 1/\tilde{t}) = 0 \qquad \text{for } \tilde{t} < 0 \qquad (B21)$$

and

$$\lim_{\tilde{t} \to -\infty} \tilde{p}(\tilde{t}, \tilde{y}) = \tilde{y} \qquad \text{for } \tilde{y} \geq 0, \qquad (B22)$$

as $N \to \infty$.

The solution to Eqs. (B20)–(B22) can be calculated numerically. By expressing the transformed variables $\tilde{t}$ and $\tilde{y}$ in terms of more fundamental quantities, we get

$$P_{\mathrm{CC}}(N, q; n_0, n_{0*})$$
$$\approx \tilde{N}^{-1/3} \tilde{p}\left[-\frac{\tilde{N}^{2/3}}{n_{0,0*}}, \tilde{N}^{1/3}\left(1 - \frac{\alpha_1 n_0}{Nq(n_{0,0*}/N)}\right)\right] \qquad (B23)$$

and

$$P_{\mathrm{CC}}(N, q; N - n_{0*}, n_{0*})$$
$$\approx \tilde{N}^{-1/3} \tilde{p}\left[-\frac{\tilde{N}^{2/3}}{N}, \tilde{N}^{1/3}\left(1 - \frac{\alpha_1 n_0}{Nq(1)}\right)\right] \qquad (B24)$$

for large N.

If the avalanche is initiated by letting each node start from 0 with probability $q(1)$, we get

$$\langle n_{0*}^{\mathrm{i}} \rangle = N[1 - q(1)] \qquad (B25)$$

and

$$\sigma(n_{0*}^{\mathrm{i}}) = \sqrt{Nq(1)[1 - q(1)]}. \qquad (B26)$$

Provided that $q(x)$ does not depend on N, $\tilde{N}$ is fixed and the spread in $\tilde{y}$ that correspond to the initial value of m will go to zero as $N \to \infty$. Also, $\tilde{N}^{2/3}/N$ approaches zero as $N \to \infty$. In this case, the probability for an avalanche to yield complete coverage is given by

$$P_{\mathrm{CC}}(N, q) \approx \tilde{N}^{-1/3} \tilde{p}[0, \tilde{N}^{1/3}(1 - \alpha_1)], \qquad (B27)$$

for large N. Only the first two arguments to P_{CC} are kept in Eq. (B27), because the process is fully determined by N and q.

In the special case that $q(x) = x$ for all $x \in [0, 1]$, Eq. (8) yields $a = 1/n_{0,0*}$, which is a strong form of Eq. (B10). By using the same induction steps that lead from Eq. (B10) to Eq. (B11), we conclude that

$$P_{\mathrm{CC}}(N, x \mapsto x; n_0, n_{0*}) = \frac{n_{0*}}{n_{0,0*}}. \qquad (B28)$$

APPENDIX C: ASYMPTOTES FOR SPARSE PERCOLATION

Provided that the derivative of $g(x)$ is well defined at $x = 0$, we let $\lambda = g'(0)$, where $g'(x)$ denotes the derivative of $g(x)$. Then,

$$\lim_{N \to \infty} g^1_{N,n}(x) = x, \qquad \text{(C1)}$$

which means that

$$\lim_{N \to \infty} q^1_{N,N-n}(x) = x \qquad \text{(C2)}$$

and

$$\lim_{N \to \infty} P^1_{\text{cc}}(n, \ell; N) = \frac{\ell}{n} \qquad \text{(C3)}$$

according to Eq. (B28). Thus, the large N limit of Eq. (57) is

$$\lim_{N \to \infty} P_{\text{n},N}(n) = \rho(n)\tau(n,\ell)\lambda^{n-\ell}e^{n(1-\lambda)}\frac{\ell}{n} \qquad \text{(C4)}$$

$$= \frac{\ell(n\lambda)^{n-\ell}}{n(n-\ell)!}e^{-n\lambda}. \qquad \text{(C5)}$$

For large N, Eq. (C5) yields

$$\lim_{N \to \infty} P_{\text{n},N}(n) \approx \frac{\ell}{\sqrt{2\pi}}e^{n(1-\lambda)}\lambda^{n-\ell}n^{-3/2}. \qquad \text{(C6)}$$

Due to the correspondence to well investigated branching processes, Eq. (C5) is not a new result. For the special case of $\ell = 1$, Eq. (C5) is given explicitly in Ref. [32], and the general form of Eq. (C5) can easily be obtained by the theorem presented in Ref. [38].

APPENDIX D: ASYMPTOTES FOR EXHAUSTIVE PERCOLATION

In analogy with our investigation of SP, we assume that $q(x)$ has a well-defined derivative at $x = 0$ and let $\lambda = q'(0)$. For EP to be likely in the large N limit, it is required that $q(x) \leq x$ for all x, meaning that $\lambda \leq 1$. The large N behavior of Eq. (64) is partly explained by

$$\lim_{N \to \infty} \frac{\rho(u)\rho(N-u)}{\rho(N)}[Q(u/N)]^N = \rho(u)\lambda^u e^{u(1-\lambda)}, \quad \text{(D1)}$$

but it remains to investigate the role of $P^1_{\text{cc}}(N - u, 0; N)$. To this end, we consider the ratio $P^1_{\text{cc}}(N - u, 0; N)/P_{\text{cc}}(N)$. [Here, we have dropped the argument q from $P_{\text{cc}}(N, q)$.]

When $N \to \infty$, there are two processes that influence on this ratio: $q^1_{N,u}$ approaches q and N increases. The increase of N makes the involved probabilities more sensitive for the shrinking differences between $q^1_{N,u}$ and q.

Thus, there are two competing processes as $N \to \infty$. The sensitivity with respect to q is limited by the variance of the number of nodes with initial state 0*, because this variance can be seen as a rescaling of q. The change in $q(x)$ by such a rescaling scales like $q(x)/\sqrt{N}$ for large N. If $q(x)$ has a well-defined nonzero derivative at $x = 0$, the difference $q^1_{N,u}(x) - q(x)$ scales like $q(x)/N$ for large N. Hence, the decrease in the difference between $q^1_{N,u}$ and q dominates over the increase in sensitivity, meaning that

$$\lim_{N \to \infty} \frac{P^1_{\text{cc}}(N - u, 0; N)}{P_{\text{cc}}(N)} = 1. \qquad \text{(D2)}$$

Thus,

$$\lim_{N \to \infty} \frac{P_{\text{u},N}(u)}{P_{\text{cc}}(N)} = \rho(u)\lambda^u e^{u(1-\lambda)} \qquad \text{(D3)}$$

$$= \frac{(u\lambda)^u}{u!}e^{-u\lambda}, \qquad \text{(D4)}$$

where $(u\lambda)^u$ should be interpreted with the convention that $0^0 = 1$ in order to handle the case $u = 0$ properly.

1. Limit distributions for supercritical EP

If $0 < \lambda < 1$ and $x = 0$ is the only solution to $q(x) = x$ in the interval $0 \leq x \leq 1$, the exponential decay of $[Q(u/N)]^N$, in Eq. (64), with increasing u ensures that

$$\sum_{u=0}^{\infty} \lim_{N \to \infty} P_{\text{u},N}(u) = 1 \qquad \text{(D5)}$$

and

$$\left[\lim_{N \to \infty} P_{\text{cc}}(N)\right]\sum_{u=0}^{\infty} \frac{(u\lambda)^u}{u!}e^{-u\lambda} = 1. \qquad \text{(D6)}$$

Thus, $\lim_{N \to \infty}$ has a unique value for each λ. This value can be calculated by considering the simplest case, $q(x) = \lambda x$. From the definition of the spreading process, we get

$$P_{\text{cc}}(N, x \mapsto \lambda x; n_0, n_{0*}) = P_{\text{cc}}(N, x \mapsto x; n_0, n_{0*}). \qquad \text{(D7)}$$

Then, Eq. (B28) and averaging over initial configurations yield

$$P_{\text{cc}}(N, x \mapsto \lambda x) = 1 - \lambda, \qquad \text{(D8)}$$

which means that

$$\lim_{N \to \infty} P_{\text{cc}}(N, q) = 1 - \lambda \qquad \text{(D9)}$$

for all q that satisfy the above mentioned criteria. We get,

$$\lim_{N \to \infty} P_{\text{u},N}(u) = (1 - \lambda)\frac{(u\lambda)^u}{u!}e^{-u\lambda}, \qquad \text{(D10)}$$

which for large u means that

$$\lim_{N \to \infty} P_{\text{u},N}(u) \approx \frac{1 - \lambda}{\sqrt{2\pi}}e^{u(1-\lambda)}\lambda^u u^{-1/2}. \qquad \text{(D11)}$$

2. Scaling at the EP transition

This section aims to derive the asymptotic distribution of u for large N for critical EP with $\alpha_2 > 0$ and $\ell = 0$. Define α_1^1, α_2^1 and $\tilde{N}^1$ analogous to the definitions of α_1, α_2 and $\tilde{N}$ in Eqs. (41) and (B7). The derivatives of $q^1(x)$ in Eq. (50) at $x = 0$ are given by

$$(q^1)'(0) = q'(u/N)\frac{1 - u/N}{1 - q(u/N)} \tag{D12}$$

and

$$(q^1)''(0) = q''(u/N)\frac{(1 - u/N)^2}{1 - q(u/N)} \tag{D13}$$

and we get

$$\tilde{N}^1 = \tilde{N} + \mathcal{O}(u/N) \tag{D14}$$

and

$$\alpha_1^1 = \alpha_1 + (\alpha_1 - 1 - 2\alpha_2)u/N + \mathcal{O}(u^2/N^2). \tag{D15}$$

For critical networks, with $\alpha_1 = 1$, we get $\tilde{N} = N/\alpha_2$ and

$$\alpha_1^1 = 1 - 2u/\tilde{N} + \mathcal{O}(u^2/N^2). \tag{D16}$$

Insertion into Eq. (B27) yields

$$P_{\mathrm{cc}}^1(N - u, 0; N) \approx \tilde{N}^{-1/3}\tilde{p}(0, 2\tilde{N}^{-2/3}u) \tag{D17}$$

for $u \ll N$.

Because $Q(x) \leq 1$ with equality if and only if $q(x) = x$, $P_{\mathrm{u},N}$ that vanishes exponentially, as N goes to infinity, for any fixed u/N such that $q(u/N) \neq u/N$. For a typical network with $\alpha_2 > 0$ at the EP transition, the only solution to $q(x) = x$ is $x = 0$. For such a network, the large N behavior of $P_{\mathrm{u},N}$ is found by expanding $Q(x)$ around $x = 0$. To the leading non-vanishing order, we get

$$Q(x) \approx \exp\left(-\frac{\alpha_2^2 x^3}{2}\right), \tag{D18}$$

which yields

$$P_{\mathrm{u},N}(u) \approx \rho(u)\exp\left(-\frac{u^3}{2\tilde{N}^2}\right)P_{\mathrm{cc}}^1. \tag{D19}$$

Hence,

$$P_{\mathrm{u},N}(u) \approx \tilde{N}^{-1/3}\rho(u)\exp\left(-\frac{u^3}{2\tilde{N}^2}\right)\tilde{p}(0, 2\tilde{N}^{-2/3}u), \tag{D20}$$

with asymptotic equality for large N. The probability density, $P(\tilde{u})$, for the distribution of $\tilde{u} = \tilde{N}^{-2/3}u$ as $N \to \infty$ approaches

$$P(\tilde{u}) = \frac{\exp(-\frac{1}{2}\tilde{u}^3)}{\sqrt{2\pi\tilde{u}}}\,\tilde{p}(0, 2\tilde{u}). \tag{D21}$$

APPENDIX E: EXACT RESULTS

A network with $g(x) = x$ is critical for all x. For such a network, $G(x) = 1$ for all x and $P_{\mathrm{cc}}^1(n, \ell; N) = \ell/n$. Hence, Eq. (57) yields

$$P_{\mathrm{n},N}(n) = \frac{\rho(n)\rho(N - n)}{\rho(N)}\frac{\tau(n, \ell)}{\tau(N, \ell)}\frac{\ell}{n} \tag{E1}$$

$$= \frac{\ell}{n}\binom{N - \ell}{n - \ell}\frac{n^{n-\ell}(N - n)^{N-n}}{N^{N-\ell}}. \tag{E2}$$

For n and N satisfying $n \gg 1$ and $N - n \gg 1$, we get

$$P_{\mathrm{n},N}(n) \approx \frac{\ell}{\sqrt{2\pi}}\frac{\sqrt{N}}{n\sqrt{n(N - n)}}. \tag{E3}$$

In the special case $\ell = 1$, $P_{\mathrm{n},N}(n)$ is the distribution of the number of predecessors to an element in a random map. This distribution, which is consistent with eq. (E2) for $\ell = 1$, was obtained in Ref. [37] and restated in Ref. [36].

For completeness, we provide an explicit expression for the distribution of avalanche sizes in the case that $g(x)$ is a first order polynomial. From Eq. (57), we get

$$P_{\mathrm{n},N}(n) = P_{\mathrm{u},N}(u)$$

$$= \frac{n^{n-\ell}u^u}{N^{N-\ell}}\binom{N - \ell}{n - \ell}\left(g(0) + \frac{\ell}{n}\right)$$

$$\times \left(\frac{ug(0) + ng(1)}{n}\right)^{n-\ell}\left(\frac{nq(0) + uq(1)}{u}\right)^u. \tag{E4}$$